

STATISTICAL AND MATHEMATICAL APPROACHES  
IN HIV/AIDS MODELING: A REVIEW

Steven J. Schwager  
Biometrics Unit  
337 Warren Hall  
Cornell University  
Ithaca, NY 14853-7801  
U.S.A.

Carlos Castillo-Chavez  
Biometrics Unit/Center for Applied Math.  
341 Warren Hall  
Cornell University  
Ithaca, NY 14853-7801  
U.S.A.

Herbert Hethcote  
Dept. of Mathematics  
University of Iowa  
Iowa City, Iowa 52242  
U.S.A.

BU-1059-M

September 1989

Abstract

This paper provides a brief introduction to the field and the literature of statistical and mathematical approaches to HIV/AIDS modeling. The emphasis is on succinct description of the methodology and on discussion of the work of some researchers who were unable to contribute to this volume. This review is not complete; however, it is complemented to a great degree by the partial reviews of the literature provided by several authors in their contributions to this volume. An additional objective of this introductory chapter is to describe to nonspecialists, in general terms, some of the main statistical and mathematical approaches currently used in problems related to AIDS epidemiology.

STATISTICAL AND MATHEMATICAL APPROACHES  
IN HIV/AIDS MODELING: A REVIEW

Steven J. Schwager  
Biometrics Unit  
337 Warren Hall  
Cornell University  
Ithaca, NY 14853-7801  
U.S.A.

Carlos Castillo-Chavez  
Biometrics Unit/Center for Applied Math.  
341 Warren Hall  
Cornell University  
Ithaca, NY 14853-7801  
U.S.A.

Herbert Hethcote  
Dept. of Mathematics  
University of Iowa  
Iowa City, Iowa 52242  
U.S.A.

Abstract

This paper provides a brief introduction to the field and the literature of statistical and mathematical approaches to HIV/AIDS modeling. The emphasis is on succinct description of the methodology and on discussion of the work of some researchers who were unable to contribute to this volume. This review is not complete; however, it is complemented to a great degree by the partial reviews of the literature provided by several authors in their contributions to this volume. An additional objective of this introductory chapter is to describe to nonspecialists, in general terms, some of the main statistical and mathematical approaches currently used in problems related to AIDS epidemiology.

1. Introduction

Mathematical models are an important tool in the study of the transmission of human immunodeficiency virus (HIV), the etiological agent for acquired immune deficiency syndrome (AIDS). These models are useful in furthering our understanding of many aspects of HIV dynamics such as short- and long-term prediction of HIV (and AIDS) incidence, the effects of behavioral changes and preventive measures on its transmission dynamics, its effects on demographic parameters, and the results of its direct interaction with the immune system. Reviews of mathematical models for the transmission of HIV and AIDS were given by Isham (1988) and Anderson (1988b). Statistical issues arise in the analysis of these mathematical models, especially in the drawing of inferences from real data, either in the generation of predictions or in the estimation of key parameters. It is thus natural that statistical

methodology plays an important role in research on HIV/AIDS. An annotated bibliography of quantitative methodology related to HIV/AIDS was given by Fusaro *et al.* (1989). It emphasized mathematical and statistical models of the magnitude of the AIDS epidemic, and statistical techniques for estimating key aspects of the disease process.

Statistics attempts to discover the underlying structure from observed data in the presence of random variability. Statistical techniques have been developed to accomplish several tasks addressing this objective. Estimation is perhaps the best known of these. There are two standard approaches, parametric and nonparametric estimation.

Parametric estimation proceeds from the assumption that the distributional form of a randomly varying quantity, or random variable, is known. A particular distribution of this form is determined by the value of a parameter, consisting of one or more quantities associated with the distributional form. For example, an incubation time may be known (or assumed) to have a gamma distribution, with probability density function  $f(x; r, \lambda) = \lambda^r \Gamma(r)^{-1} x^{r-1} e^{-\lambda x}$  for  $x > 0$ ; the parameter is the vector  $(r, \lambda)$ , which completely determines one member of this family. Estimation procedures seek to determine a single value (point estimate), or a range of values (confidence interval or region), representing our best judgment about the true value of the unknown parameter. Several standard methods of estimation have been developed, including maximum likelihood and least squares, each with properties that make it a good choice under certain circumstances. Maximum likelihood estimators, or MLEs, are the most commonly used because of their desirable theoretical properties: under broad regularity conditions, MLEs are consistent, asymptotically normal, and asymptotically efficient. It may be that only some, but not all, of the components of the parameter are of primary importance; the rest are then called nuisance parameters. Another possibility is that a function of the parameters, such as  $r/\lambda$ , is of special interest, perhaps because of its epidemiological interpretation. Both of these situations can be handled by appropriate estimation techniques. A general treatment of parametric estimation is given by Mood, Graybill, and Boes (1974); a fuller treatment of parametric point estimation is given by Lehmann (1983).

Nonparametric estimation, in contrast, seeks to obtain results when the distributional form of a random variable is not known. Nonparametric estimation methods must therefore perform well over a broad range of distributional forms. For example, we may want to estimate the mean of a distribution whose form cannot be specified because we lack sufficient information. This is inherently more difficult than parametric estimation because of the greater generality of the problem: many families of distributional forms may be consistent with the observed data, and nonparametric estimates must be appropriate for this whole range of families. Nonparametric estimation is treated by Conover (1980) and Lehmann (1975).

There is an intermediate approach, semiparametric estimation, in which the unknowns include both parameters and functions. The parameters in the model are usually of primary interest, with the unknown function(s) serving as extremely flexible nuisance parameter(s). The proportional hazards model (Cox, 1972) is a semiparametric model useful in AIDS and other epidemiological research.

It is important to distinguish between estimation of a parameter or other quantity associated with a distribution and estimation of the distribution itself. For example, the distribution of incubation times or reporting lags of AIDS cases may carry important information not retained by the mean and variance of this distribution. Distributions can be estimated by both parametric and nonparametric approaches. Parametric estimation of the distribution is closely related to testing distributional goodness of fit. This will be discussed shortly. Nonparametric density estimation techniques have developed along several lines. The most commonly used is the smoothing kernel approach, which spreads the atom of probability mass of each observation continuously across a neighborhood of its location. Splines and other methods have also been used. Details can be found in Silverman (1986) and Eubank (1988).

The testing of hypotheses is another major statistical activity related to mathematical modeling, as it seeks to determine whether observed data are consistent with proposed hypotheses about the underlying mechanism. In the usual testing formulation, we specify a null hypothesis  $H_0$  of "no difference" or "no effect" and an opposing alternative hypothesis  $H_1$ . Then we compute a test statistic that quantifies the strength of the evidence in the observed data against the null hypothesis. For example, a null hypothesis might claim that the mean length of the incubation time does not differ for various routes of exposure to HIV; the alternative hypothesis might be very broad, e.g., that there are some differences among routes, or more narrow, e.g., that routes of exposure differ in a highly specified pattern (mean incubation time — latent period plus asymptomatic infectious period — is greater for sexual contact than for blood transfusion, greater for blood transfusion than for IV drug use and so on), as suggested by the researcher's knowledge of the process. Statistical methodology for hypothesis testing is well developed, although there are still many specialized questions that need further investigation. For details, see Mood, Graybill, and Boes (1974) and Lehmann (1986).

An important class of hypothesis tests deals with the question of whether or not an observed set of data is consistent with a specified distribution. This class of procedures, called goodness-of-fit tests, can be illustrated by considering a test of the null hypothesis  $H_0$ : data came from a gamma population with  $(r, \lambda)$  unknown. The alternative hypothesis may be either very general, e.g.,  $H_1$ : data came from other than a gamma population, or more specific, e.g.,  $H_1$ : data came from a contaminated gamma population, or  $H_1$ : data came from

a mixture of gamma distributions. Some or all of the parameter values may be known in either the null or the alternative hypothesis, or both.

Predictions of the future course of the AIDS epidemic provide important information for public policy formulation. Anderson (1988a) lists several key questions for government and health planners, i.e., the likely magnitude of the epidemic and the level of change in the population's sexual behavior needed to affect this magnitude. Various aspects of the disease are of interest, i.e., the prevalence and incidence of HIV infection, the number of AIDS cases, and the number of deaths of patients diagnosed as having AIDS. We need estimates of the future values of these and other aspects of the epidemic. With sufficient data, short-term predictions for 1-3 years into the future can be made with reasonable accuracy. Prediction over a longer period (using statistical approaches) appears to be much more difficult owing to the greater uncertainties involved. At present, transmission models may provide us with the best approach to the generation of mid- to long-term predictions. Three approaches have been applied to prediction of various features of the course of HIV and AIDS: direct extrapolation of the recent data by empirical curve fitting; the back-calculation approach, which combines AIDS incidence data with an estimate of the incubation period to obtain an estimate of the pattern of HIV seroincidence over time, leading to extrapolations for future levels of HIV seroprevalence and AIDS incidence; and dynamic transmission models, based on systems of differential or difference equations quantifying the rates of transition between disease states, including susceptible, infected by HIV, and diagnosed as having AIDS. These will be discussed in detail below. Many aspects of short-term prediction of HIV infection and AIDS are treated in the report of the British Working Group chaired by Cox (Working Group 1988, see Chapters 3, 4, Appendices 3, 6).

The main body of this paper is organized as follows. Section 2 describes some special difficulties that occur in the analysis of AIDS data. Section 3 discusses the short-term prediction of various aspects of the HIV/AIDS epidemic. Section 4 treats the estimation of epidemiological parameters and distributions. Section 5 briefly reviews some of the literature on dynamic models for HIV/AIDS. We do not review the literature on models for the immune system; however, relevant references can be found in the articles by Perelson (1989), Merrill (1989), Hoffman and Grant (1989), and Layne *et al.* (1989) in this volume, and in the recent work of Anderson and May (1989).

## 2. Special problems of AIDS data analysis

Several special problems must be dealt with in the analysis of data on HIV infection and AIDS. These are rooted in the current state of knowledge about the disease and its

modes of transmission. Many aspects of the biological mechanism underlying the disease and of the social and behavioral processes through which HIV and AIDS are spread are not well understood. As can be seen from the articles in this volume, researchers have begun to identify these mechanisms through the use of dynamical systems. Nevertheless, until further progress is made in these areas, statistical methodology must be applied in the presence of fundamental uncertainties, making analysis considerably more difficult. Not surprisingly, the interplay between dynamic models of disease transmission and statistical analysis is playing a leading role in elucidating the unknown aspects of the disease and its transmission mechanisms.

There are substantial difficulties related to the availability and quality of AIDS data suitable for statistical analysis. The report of a workshop on modeling AIDS epidemiology (Workshop 1988, Sec. 4) discussed data availability and quality and made recommendations about meeting the data needs of AIDS research. A statement by the Royal Statistical Society (1988) on data requirements and collection addressed the same issues. Layne *et al.* (1988) suggested that governments should establish national HIV databases for the use of researchers and public health officials. Data come from sample surveys which seek to determine specified features of a chosen population and from the operation of health organizations such as hospitals. The population of interest may be national, regional, or local, and the sample may range from a carefully designed probability sample to a convenience sample or a group of volunteers. The workshop report noted the need for national representative data on HIV prevalence, HIV and AIDS incidence, and prevalence of risk behaviors and size of risk groups. The collection of such data is complicated by numerous problems that make many important epidemiological parameters extremely difficult to estimate. Several of these problems will now be mentioned; the need to develop methods for overcoming them is urgent.

Parameters like the average incubation period of AIDS, duration of infectiousness, and time to death of AIDS patients involve long periods of time. The lengths of some of these periods are much greater than can be satisfactorily estimated from the eight years of observational data currently available. (State and local health officials in the U.S. have been reporting AIDS cases to the CDC since 1981, and the definition of AIDS has changed over the course of the epidemic.) Until data have accumulated over a longer interval, providing observations close to the upper extremes of the incubation period and other periods of interest, analysis will inevitably be incomplete and inaccurate. Medley *et al.* (1987, 1988), using data on transfusion-associated cases of AIDS in the U.S., noted the strong dependence of the estimated average incubation period of AIDS on the parametric model chosen and judged their work on characterizing the incubation period as "tentative until more of the incubation period is observed."

Reporting of AIDS cases often involves a time lag between the diagnosis and the receipt of the report by the central public health agency (the Centers for Disease Control in the U.S.A., the Communicable Disease Surveillance Centre in the United Kingdom, etc.). This lag can range from a few weeks to 1-2 years. Anderson (1988b, p.242) described the lag as "increasing and variable," and Harris (1987) said that the distribution of delays has been shifting to the right, with median delay up 0.6 months from mid-1986 to March 1987. The British report, however, hedges (Working Group 1988, p.31), and its Appendix 8 tends weakly to support the opposite view.

Underreporting of AIDS cases is also a major concern. This may stem either from failure to diagnose a case or from failure to report a case that has been diagnosed. One aspect of this is self-selection bias in the medical observation process: many patients who are at high risk choose to avoid diagnosis and treatment, and thus remain undetected. This often occurs even when HIV antibody testing has been completed; see McCusker *et al.* (1988) and Lyter *et al.* (1987).

A related issue is the presence of cases that do not satisfy the clinical criteria of AIDS (which have evolved over time) but whose condition is influenced by HIV infection. Some individuals may not develop "full-blown" AIDS but may nevertheless die as a result of HIV infection, perhaps indirectly; some physicians have reported seeing about three times as many cases of non-AIDS HIV-associated illness as cases of AIDS (Working Group 1988, p.37).

Incomplete data can arise in numerous ways. There can be a failure to collect information that subsequently turns out to be necessary for analysis. The attempt to collect information can result in missing values, for example, laboratory measurements not made or questions not answered by the patient. Patients can (and often do) drop out of clinical trials and other observational studies prematurely.

The quality of data can be adversely affected by biases of several kinds. The self-selection bias described above and other refusals by patients at risk for HIV infection and AIDS to provide information are forms of nonresponse bias. Another possibility is the misreporting, either accidentally or intentionally, of medical and behavioral information. This response bias has many causes, ranging from the phrasing of questions (Schuman and Presser 1981) to some respondents' selecting answers to gain social approval (Bradburn, 1983). Fox and Tracy (1984) cited many examples of both response and nonresponse bias on sensitive questions.

The behavioral and sociological processes related to the spread of AIDS must be addressed. Some work of this kind has been done. Baldwin and Baldwin (1988) studied factors associated with AIDS-related sexual high-risk behavior among Southern California college students. McCusker *et al.* (1988) reported that HIV antibody testing reduced the level of sexual activity of a group of homosexuals in Boston. Johnson (1988) reviewed

research on sexual and drug-use behavior in the population and considered future research on the social determinants and prevention of HIV transmission. Kaplan *et al.* (1987) surveyed the sociological literature on AIDS and suggested a research agenda based on models for the social factors related to the onset and the course of AIDS. In this volume, Aron and Sarma (1989) use behavioral data from a 1988 U.S. survey on sexual partners to examine the risk of HIV spread in the population.

The profile of HIV infection and AIDS incidence is changing over time. Anderson (1988b, p.242) cited "the changing shape of the epidemic as saturation effects occur and the infection moves from high-risk group into the general population" as one of the two major problems with exponential extrapolation. The behavioral patterns of high-risk activity in the population are evolving, partially in reaction to the epidemic. Estimation under these conditions is an attempt to hit a moving, rather than a stationary, target.

A final complicating factor is the recent modification in the clinical definition of AIDS. For full details, see Centers for Disease Control (1987); Karon *et al.* (1988) summarized the changes. The classification of a patient as having or not having AIDS is consequently time-dependent; two identical patient histories, one before and one after the change, could result in different diagnoses. This problem affects many aspects of the statistical analysis of data on HIV infection and AIDS.

If mathematical models of the AIDS epidemic are to provide reliable results, they must factor in the limitations of the data in sample size and accuracy of observed values. In this volume, De Gruttola and Lagakos (1989b) examine how these limitations can affect the uncertainty of estimation using models.

### 3. Short-term Prediction of HIV Infection and AIDS

Three approaches that have been applied to forecasting the course of the AIDS epidemic are direct extrapolation (see Morgan and Curran 1986; Karon *et al.* 1989; and references therein), back-calculation (see Brookmeyer and Gail 1986, 1988; Brookmeyer and Damiano 1989; Colgate *et al.* 1989; and references therein), and dynamic transmission models (see Anderson 1988b and references therein). These have been used to predict the prevalence of HIV infection, the incidence of new cases of AIDS, and the number of deaths associated with HIV and AIDS. These methods require data on the quantity of interest over at least the past few time periods. Information about a greater number of time periods allows for more accurate analysis; for example, it helps in choosing which mathematical function to use in the extrapolation approach. Data on the values of related quantities are needed in the back-calculation and dynamic transmission model approaches. Partitioning the population into



subgroups, analyzing each of these separately, and then combining the resulting predictions into an overall population prediction is often effective when the available data make this possible.

Prediction involves four broad kinds of errors: model errors, caused by inadequacy of the model in representing the actual process; shift errors, caused by future changes in the nature of the process; estimation errors, caused by the estimation of the unknown parameters in a specified prediction model; and random errors or statistical fluctuations, caused by the variability present in any stochastic process even when model and estimation errors do not occur (Gilchrist 1983). The transmission models described in Section 5 provide a framework that can be used to generate short- and mid-term predictions for HIV and AIDS seroincidence, and therefore will not be discussed in this section. We proceed with the description of the extrapolation and back-calculation approaches.

### 3.1. Direct Extrapolation

The direct extrapolation approach involves extending a chosen mathematical function into the future. It is based on the assumption that the observed trend will continue into the future. A specified family of functions consistent with the observed data is selected, for example, the exponential ( $a \exp(bt)$ ), the logistic ( $1/[1 + a \exp(-bt)]$ ), or the damped exponential ( $a \exp(bt^c)$ ). The parameters are estimated from past data, usually by the method of weighted nonlinear least squares, and the resulting curve provides estimated future values. These estimated values may be modified to reflect reporting delays, underreporting, and other difficulties mentioned earlier. A commonly used technique for obtaining overall population estimates is to divide the population into subgroups, analyze each separately, and amalgamate the results. This kind of extrapolation, or trend analysis, gives reasonably accurate estimates for 1-3 years into the future when based on sufficient data. Such estimates are critical to our understanding of the course of the disease in the short run, since it is on this basis that current policy decisions must be made.

Extrapolation is used by the CDC for short-term projections of AIDS case incidence in the U.S. Details were given by Karon *et al.* (1988) who estimated that 365,000 AIDS cases will be diagnosed in the U.S. by 1992, with a 68% prediction interval of 205,000 to 440,000 cases. Healy and Tillett (1988) extrapolated monthly figures to the end of 1986 to estimate the number of AIDS cases in the U.K. two years into the future using several different model assumptions. Both normal and Poisson error models resulted in a roughly exponential rate of increase, with doubling time of about 14 months having remained nearly constant since mid-1985. Downs *et al.* (1987) fitted exponential functions to European AIDS surveillance data

after adjusting reported case numbers through June 1986 for the estimated delays between diagnosis and report (but not for under-reporting). They estimated a wide range of doubling times (4.3 to 17.7 months) and predicted numbers of AIDS cases through mid-1988 for various European countries. In Morgan and Curran (1986), a quadratic polynomial was fitted, using weighted linear regression, to adjusted AIDS case counts transformed by the modified Box-Cox method. The resulting polynomial was extrapolated to 1991. In this volume, Karon *et al.* (1989) base their projections on the Box-Cox procedure, modeling the proportions of AIDS cases in population subgroups over time.

The extrapolation approach, although relatively straightforward to apply, has several important drawbacks; some are related to its simplicity. It is unlikely to give realistic results beyond a few years into the future, since it takes into account neither the changing pattern of the epidemic as high-risk groups become saturated and the infection expands into the general population (Anderson, 1988b, p.242) nor changes in the behavior of the population. Moreover, the extrapolation method provides very little understanding of the epidemiological processes involved in HIV and AIDS, not being based on biological or behavioral mechanisms. Thus, it cannot be used to predict the effects of biological, behavioral, or social changes in the population or any of its subgroups. Nevertheless, it is useful insofar as it provides those who must make policy decisions with a baseline on which to plan for the immediate future.

Choosing the mathematical function to fit historical data is complicated by the agreement of many different functional forms with the data (see DeGruttola and Lagakos 1989, in this volume); these forms can produce very different predictions over even a short time period. The fit to the data can be improved by increasing the number of parameters in the chosen mathematical function, but this improved fit may actually decrease the predictive power of the model (Workshop 1989, p. 10). When subgroups within the population are considered separately and then amalgamated, lack of accurate knowledge of the sizes of high-risk subgroups, such as homosexuals, IV drug users, and prostitutes, can lead to difficulties.

In the Working Group (1988) report on short-term prediction of HIV infection and AIDS in England and Wales, direct extrapolation was used to estimate the prevalence of HIV infection and the incidence of new cases of AIDS. In the discussion of new cases of AIDS, several different mathematical functions were examined, including the exponential, the quadratic exponential, and the linear logistic. All three of these functional forms were consistent with historical data; however, this was not surprising as the abundance of parameters makes this historical agreement easy to achieve. We note, though, that the exponential produced a 1992 estimate nearly five times as great as the other two functions. The transmission models to be discussed in Section 3.3 provide us with a systematic approach for the evaluation of parametric models.

### 3.2. Back-calculation

The back-calculation, or back-projection, approach combines data on the incidence of AIDS cases with an estimate of the distribution of the length of the incubation period. The cumulative incidence of AIDS at time  $t$  is the summation up to time  $t$  of the product of the incidence of HIV at time  $\tau$  and the probability of developing AIDS within  $t - \tau$  years after infection (given that AIDS will eventually develop). If the incidence of AIDS up to time  $t$  and the distribution of the AIDS incubation period are known, it is possible to calculate the incidence of HIV up to time  $t$ . In other words, since the cumulative incidence of AIDS is a convolution, the incidence of HIV can be found by deconvolution. Therefore, the incidence of HIV can be estimated by deconvolution from data on AIDS incidence adjusted for reporting lags and underreporting, and the estimated distribution of the length of the incubation period. This leads to extrapolations for future levels of HIV infection and AIDS incidence.

Back-calculation yields predictions for homosexuals and IV drug users that are reliable for 1-3 years into the future. However, for the general population, there is more uncertainty: AIDS incidence data are currently much more limited, and the estimated distribution of the incubation period comes from studies of high risk groups, raising doubts about its validity for the general population (Workshop 1988, p. 10). The accuracy of the estimated distribution of the incubation period is critical; the back-calculation method is very sensitive to the distribution assumed for the incubation period (Brookmeyer and Gail 1986, 1988, Brookmeyer and Damiano 1989). This is especially troublesome because it is currently necessary to assume a functional form and estimate the parameters of this distribution, and data are available only for a limited range of values of  $t$ ; thus, we know much about the lower tail of the distribution of the incubation period but little about the upper tail. Back-calculation "can establish a range of plausible estimates of HIV seroprevalence, but the error range is wide and confidence limits are not available" (Workshop 1988, p. 10).

Brookmeyer and Gail (1986, 1988) used back-calculation to estimate the minimum size of the AIDS epidemic in the U.S. They used the Weibull distribution with a median of 4.3 years, obtained earlier by other authors, to estimate the incidence of HIV before 1986, and then estimated the number of AIDS cases from 1986 to 1991 in people infected before 1986. However, Hyman and Stanley (1988) found such estimates to be very sensitive to the assumed nature of the probability distribution of the length of the incubation period; considering the continuous analog of the back-calculation process, they determined that the confidence intervals on the forecasts were very wide. The sensitivity of model predictions to assumptions about the distribution of the incubation period was also noted by Anderson *et al.* (1987) and Anderson (1988a).

#### 4. Estimation of Epidemiological Parameters and Distributions

The epidemiology of HIV infection and AIDS is characterized by the distributions of various quantities associated with the disease. Some of these are intervals of time, e.g., the incubation period, others are probabilities, e.g., the probability of transmission per sexual contact and rates, e.g., the rate of change of sexual partners. A simple approach is to treat these as unknown constants, or parameters of the disease model, which we seek to estimate from observed data. A more realistic approach, though, is to view these as random quantities, each varying from individual to individual according to an unknown probability distribution. For some purposes, knowing (or estimating) some particular numerical aspect or parameter of one of these distributions will suffice, while for other purposes, knowing (or estimating) the entire probability distribution or probability density function is necessary. Whether we need to estimate quantities or distributions is closely related to the type of questions that we wish to address. These issues are intimately connected to dynamic models. (See the contributions in this volume by Castillo-Chavez *et al.* 1989a, Thieme and Castillo-Chavez 1989a; Jacquez *et al.* 1989; Koopman *et al.* 1989; Kaplan *et al.* 1989; Cardell and Kanouse 1989; or see Huang *et al.* 1989 and Thieme and Castillo-Chavez 1989b).

For example, the mean length of the AIDS incubation period is an important piece of information, so we may seek to estimate this parameter. The standard deviation of the incubation period is another parameter we may want to estimate. But these two values do not give information about the distribution of the length of the incubation period. Rather than estimating the mean and standard deviation directly, a different approach is to estimate its entire probability distribution, from which the mean, standard deviation and other parameters can be found. Lui *et al.* (1988) and Giesecke *et al.* (1988) are among those who have modeled the incubation period distribution with parametric families. In this volume, Longini *et al.* (1989) modeled the incubation period as a sum of a two-parameter (time-delayed) exponential variable and three time-homogeneous Markov process waiting times, which gives a generalized gamma distribution.

A nonparametric approach to estimating the distribution of the incubation period was taken by Lagakos *et al.* (1988). They handled the problem of right truncation of the data by transforming to reverse time, obtaining left truncated survival data. They developed nonparametric methods for estimating and comparing identifiable aspects of incubation period distributions for several population groups. De Gruttola and Lagakos (1989a) developed nonparametric methods for doubly censored survival data in which both the time of infection and the onset of AIDS can be truncated. These nonparametric methods constitute a response to the fact that the distribution of the incubation period is not identifiable outside the range of incubation times spanned by the observed (truncated) data.

The focus of the inference process is on this range. To estimate the entire distribution for use in epidemic models for the spread of AIDS and other purposes, it is necessary to specify a parametric model and to use likelihood or similar methods. Lagakos *et al.* (1988) pointed out the fundamental problem of the parametric approach through an example in which Weibull distributions with medians of 8.5 years and 210 years both agreed well with the data. This demonstrated the inherent weakness of parametric estimates of quantities related to the unobserved portion of the distribution. A comprehensive discussion of the use, strengths, and weaknesses of nonparametric approaches appears in this volume (De Gruttola and Lagakos, 1989b). We note, however, that some of the difficulties pointed out by the example in Lagakos *et al.* (1988) can be overcome through the use of dynamic models. The simplest is to feed these parametric representations for the incubation period into a dynamic model and then to require realistic predictions not only of the number of AIDS cases but also of the number of HIV infected. The theoretical basis for this approach has been developed and has been applied to fit models optimally for the dynamics of interacting populations (see Banks *et al.* 1989 and references therein).

Kalbfleisch and Lawless (1989) considered both nonparametric and parametric estimation from data on transfusion-related AIDS. The data were compiled by the Centers for Disease Control using retroactive ascertainment: when an individual is diagnosed as having AIDS, often the time of the initiating event, infection with HIV by blood transfusion can be determined exactly, or nearly so. Kalbfleisch and Lawless constructed the likelihood function of the data, noting that it had been used by Brookmeyer and Gail (1988), Lagakos *et al.* (1988), Lui *et al.* (1986), and Medley *et al.* (1987, 1988). They examined nonparametric estimation based on the occurrence of initiating events as a Poisson process. Neither the distribution of the incubation period nor the Poisson process cumulative intensity function is estimable, although time-truncated versions of them are. This identifiability problem was pointed out by Lagakos *et al.* (1988). Adopting a parametric model seems to solve the problems inherent in nonparametric estimation by making the quantities of interest functions of the parameters which can be estimated by maximum likelihood from the data; however, these functions may be estimated very poorly. Kalbfleisch and Lawless showed that this happens with the transfusion data, reflecting the fact, clearly visible in the nonparametric analysis, that these data have virtually no information on percentiles of the distribution of the incubation period, expected number of infections leading to a diagnosis of AIDS, and other quantities of interest. Consequently, the parametric model point estimates for these are imprecise.

The term "parameter" has two slightly different meanings. A probability distribution in a specified family is characterized by a vector of parameter values; the probability density function is an expression involving these parameters. For example, a normal distribution is

characterized by its mean and variance ( $\mu, \sigma^2$ ), a binomial distribution by the pair ( $n, p$ ). A considerable amount of statistical literature is devoted to methods of estimating these parameters, which we can call “statistical model parameters,” and to criteria for comparing competing estimation methods. An “epidemiological parameter,” on the other hand, is any quantity whose value is useful in understanding the epidemiology of the situation. Epidemiological parameters may or may not be statistical model parameters, and vice versa. For example, Anderson (1988b) describes models of HIV spread in different population subgroups for which the “effective average rate of partner change” should be defined as  $c = m + v/m$  where  $m$  is the mean rate of partner change and  $v$  is the variance;  $c$  is an epidemiological parameter here but probably not a statistical model parameter in the usual sense.

Estimation is often accompanied by related analyses. When estimating a vector of statistical model parameters, we are frequently interested in testing the hypothesis that some of them take on special values. These values often reduce the model to a simpler special case. For example, in modeling a Weibull distribution with a time delay, Longini *et al.* (1989) test the hypothesis that the Weibull shape parameter  $\alpha$  equals 1, reducing the distribution to an exponential. Another type of analysis related to parameter estimation is testing whether the distribution being fitted is consistent with the observed data by using a goodness-of-fit test. Still another type of analysis is testing the dependence among parameters, e.g., the effects of cofactors and markers (other STD's, age, health status, etc.) on the distributions of time spent in the stages of HIV infection. Brookmeyer *et al.* (1987) pointed out several potential sources of bias in an AIDS prevalent cohort study of cofactors and markers under a proportional hazards model. Other recent studies of cofactors and markers were reported by Goedert *et al.* (1987) and Padian *et al.* (1987). A general discussion of statistical techniques for addressing many of these issues can be found in Neter *et al.* (1985).

## 5. Dynamic Models

The first epidemiological model was presented by Daniel Bernoulli (1760) to the Royal Academy of Sciences in Paris. The mathematical theory of epidemiology made no significant advances until the work of the Russian physician P. D. En'ko was published in 1889 (see Dietz 1988a). En'ko constructed the first chain binomial model (wrongly attributed to Reed, see Dietz 1988a for further details). The so-called Reed-Frost models still play a very important role in theoretical and applied epidemiology. The key concepts in the development of the mathematical foundations of theoretical epidemiology derive from the work of the Nobel laureate Sir Ronald Ross (1911), although partial credit should be given to Brownlee

(1907) and McKendrick (1912). Ross introduced the assumption that the rate of new infections is proportional both to the number of susceptibles and to the number of infectious individuals (the so-called "mass-action law"), developed the first mathematical model for the spread of a vector-transmitted disease (malaria), and concluded that, to eradicate malaria, it was sufficient to bring the vector population below a threshold level. This theoretical result, the first threshold theorem, provided the theoretical foundation for the development of control programs. Ross used this result to conclude that a successful control program for malaria did not require the elimination of the whole mosquito population. McKendrick extended these results, and in 1927, he co-authored his celebrated paper with Kermack (see Kermack and McKendrick 1927) establishing that a threshold number of susceptible individuals must be available if an epidemic is to take place.

Ross was aware of the necessity of taking into account the effects of nonhomogeneous mixing, demography, geographical distribution and other factors in order to increase the predictive and explanatory power and applicability of epidemiological models. This level of detail could only be introduced by the stratification of a population into subpopulations according to specified criteria and by a detailed description of the mixing between subpopulations. Further elaboration of these ideas had to await the dramatic increases observed in venereal diseases during the 1960's, and was finally propelled further by the AIDS epidemic.

The first mathematical model for the transmission of a venereal disease was developed by Cooke and Yorke (1973). A model for gonorrhea with an arbitrary number of randomly (proportionally mixing) interacting groups was formulated and analyzed by Lajmanovich and Yorke (1976). Hethcote (1976), by observing that modeling sexually transmitted diseases for two-sex populations is mathematically equivalent to the modeling a host-vector interaction, took us back to the seminal work of Ross. Yorke, Hethcote and Nold (1978) and Hethcote and Yorke (1984) introduced concepts such as saturation and preemption, as well as the concept of core subpopulation into gonorrhea analysis. The idea of the core subpopulation has been extremely important in theoretical epidemiology. Its importance in disease dynamics was clearly understood by Ross and was later used by MacDonald (1957) as an explanatory mechanism for the lack of success in eliminating malaria.

Many mathematical models for HIV transmission and AIDS incidence have dealt primarily with one homogeneously mixing risk group which usually consists of highly sexually-active homosexual men. Some of these modeling efforts are described below.

## 5.1 Single-group models

Anderson *et al.* (1986) described some preliminary attempts to use mathematical models for HIV transmission in a homosexual community. The epidemic data available on HIV infection and the incidence of AIDS was surveyed. After the risk groups and transmission mechanisms were described, doubling times for AIDS incidence were given for risk groups in various geographic locations. Some data were also given for the HIV infectious period, the proportion who develop AIDS, and measures of sexual activity. Models of the early stages of the AIDS epidemic in homosexual men were used to find the reproductive number from the distribution of the AIDS incubation period and the initial doubling time. Their more complex models showed that heterogeneity in sexual behavior can greatly influence the predictions, with more heterogeneity implying decreased magnitude of the AIDS epidemic. This result is reasonable since high heterogeneity implies that the few very sexually active people are removed rapidly from the infectious pool. Anderson *et al.* (1986) emphasized that uncertainty in parameter values implies that the models are not suitable for prediction. The purpose of their modeling was to investigate the effects of various parameters and to help improve our general understanding of the transmission dynamics of HIV infection. Areas of biological uncertainty, future data needs, and public health policy implications were discussed.

Pickering *et al.* (1986) formulated a model for the spread of HIV and AIDS incidence in the homosexual male population in three large cities. They used a discrete time nonlinear model for the sexual transmission of HIV with several possible courses of progression after infection. The models used trends in anal-rectal gonorrhea incidence to determine the changes in homosexual behavior. They gave some preliminary forecasts for San Francisco, Los Angeles and New York City but concluded that there were insufficient data to choose between radically different forecasts.

May and Anderson (1987) presented some simple HIV transmission models to help clarify the effects of various factors on the overall pattern of the AIDS epidemic. They began by defining the basic reproductive number as the product of three parameters and then obtained estimates of these three parameters from various data sources. They showed that if the probability of developing AIDS increases linearly with time since infection, then the distribution of the AIDS incubation period is a Weibull distribution. Their calculations assumed that 30% of HIV infecteds eventually develop AIDS, but we now know that this percentage is too low. They considered a model for heterosexual transmission where infection comes from the homosexual male population through bisexuals and found that the doubling times would be significantly larger in the heterosexual population than in the homosexual population. At present, this is not a realistic model for the sexual transmission of HIV in the



United States, since most heterosexual transmission is to sexual partners of intravenous drug users. In their discussion, they emphasized the uncertainty of the parameter values and the need for better data in several areas.

Blythe and Anderson (1988a) considered HIV transmission models with four forms for the distribution of AIDS incubation period (exponential, Weibull, gamma, and rectangular). As in most models, the HIV infectious period was assumed to be equal to the AIDS incubation period. The impact of the four distributions on HIV transmission dynamics in male homosexual communities was assessed by examining the equilibrium states and their local stability in a model with constant recruitment of susceptibles. In their discussion of the relative merits of the four distributions of the AIDS incubation period, they concluded that, for qualitative purposes, it may be sufficient to consider only these four distributions (if their means coincided with the observed value). Castillo-Chavez *et al.* (1989a, b, c) have extended the above results to arbitrary distributions and have analyzed a model where the mean rate of acquisition of new partners depends on the size of the sexually active population. Their results are further described in this volume (see Castillo-Chavez *et al.* 1989d).

Blythe and Anderson (1988b) also considered an HIV transmission model that encapsulates temporal variation in the infectiousness of HIV-infected persons and variability in the incubation period for AIDS. Variable infectivity was modeled in two ways for a homogeneously mixing homosexual population. They found their first approach, based on a multi-stage classification, more useful at present. Using two infectivity peaks (one after a short latent period and the other before the onset of "full-blown" AIDS), they observed that the initial phase of the infectiousness will tend to drive the early doubling time of the epidemic, while both peaks will determine the overall magnitude of the epidemic and the magnitude of the unique endemic state. Thieme and Castillo-Chavez (1989b) have carried out the theoretical analysis of a more general model that includes Blythe and Anderson's and have shown that the exclusive existence of a single initial peak could potentially force the incidence to oscillate. These authors, however, have not yet performed numerical simulations to investigate the possibility of this type of fluctuations for realistic parameter values. A discussion of their results is included in this volume (see Thieme and Castillo-Chavez, 1989a).

Bailey (1989) presented a model for HIV infection and AIDS in which infected people proceed through a sequence of stages to AIDS and then to death. The model is given by a system of  $m+2$  nonlinear differential equations with mass-action incidence term and negative exponential waiting times in the infected stages, which correspond to a gamma distribution for the AIDS incubation period. He used data on HIV prevalence in the San Francisco City Clinic cohort of 7,000 people and the reported AIDS incidence in all San Francisco and obtained a best (minimum chi-square) fit of his model. The best fit yielded a gamma distribution with  $m = 7$  for the AIDS incubation period.

A further class of models that explicitly considers the dynamics of pairs of individuals and the duration of these partnerships has been developed by Dietz (1988b) and Dietz and Haderler (1988). Their approach has consisted of the superposing of an epidemic process on the demographic "marriage" models of Kendall (1949) and Fredrickson (1971). Dietz, Haderler, and their collaborators have further clarified the process of pair-formation. Details of their most recent work can be found in Haderler (1987), Haderler *et al.* (1988), and Haderler (1989). In addition, an article that reviews this important work with a higher degree of detail and presents some new results is being included in this volume (see Waldstätter 1989).

Mode *et al.* (1988) considered a stochastic population model of an AIDS epidemic in a population of male homosexuals. Computer intensive methods were used to study some properties of the model statistically. A numerical factorial experiment was used to study three factors of importance in the evaluation of the AIDS epidemic. These factors were the distribution of the latent period of HIV, the probability of infection with HIV per sexual contact with an infected individual, and the distribution of the number of contacts per sexual partner per month. They found that the latent period of HIV infection had a decisive impact, but the impact depended crucially on the other factors. Their Monte Carlo experiment showed that the deterministic, nonlinear difference equations using expected values gave more pessimistic predictions than the stochastic population process. Their latent period of HIV would more properly be called the incubation period for AIDS. They used the Weibull and gamma distributions for this AIDS incubation period. The infectivity of HIV-positive individuals was taken to be constant and then zero when they developed AIDS. Since a longer median AIDS incubation period implies a longer infectious period, their conclusion that the HIV prevalence is much higher for longer median AIDS incubation period seems reasonable.

Tan (1989) used a stochastic model for the spread of the AIDS virus in a homosexual population. In his model, susceptible (S) persons become HIV latent (L), infective (I) and then develop AIDS (A). Transitions between these groups were governed by probabilities with contact rate and two transmission rates. The probability generating functions (PGF) of the number of latent persons infective persons; and AIDS cases were derived. The expected numbers, and variances and covariances of these persons satisfy some ordinary differential equations. These equations are solved numerically to assess the effects of various factors on AIDS spread.

Kaplan (1989) has developed dynamic models that apply to needle sharing populations. He has made the assumption that a susceptible individual using an infected needle removes the virus from the needle. Kaplan has done extensive simulations illustrating the sensitivity of the model to various parameters and has computed the basic reproductive number for his model. He has also discussed the effect of possible intervention strategies.

## 5.2 Multigroup models

The AIDS models described above have involved only one population. Clearly, HIV transmission takes place in populations that are heterogeneous in a variety of ways. The contacts between people can be homosexual, heterosexual, or by needle sharing among intravenous drug users; some groups have higher contact rates than others; people may have contacts primarily with others who are similar or with a wide variety of partners; and behavior is not uniform geographically or temporally. One way in which this heterogeneity can be modeled is to consider models with multiple groups. Another possibility is to use continuous distributions of behaviors instead of discrete groups with different behaviors. Some recent models of these types for sexually transmitted diseases and AIDS will now be described.

One of the first multigroup models for a sexually transmitted disease was the gonorrhea model of Lajmanovich and Yorke (1976). In this model with  $n$  groups, each group was divided into two classes: the susceptibles and the infecteds. Since people are susceptible again after they are cured of a gonococcal infection, gonorrhea is called an SIS disease. Their model was a system of  $n$  nonlinear ordinary differential equations where the incidences were quadratic mass action laws and the removal rates corresponded to negative-exponential distributed infectious periods. Lajmanovich and Yorke determined a theoretical threshold condition for this model. They proved global asymptotic stability of the trivial equilibrium point below the threshold. No attempt was made to estimate parameters or to apply the model.

Nold (1980) proved similar results for a generalization of the Lajmanovich and Yorke SIS model. Similar results for analogous SIR model were proved in Hethcote (1978), Hethcote and Thieme (1985), and Hethcote and Van Ark (1987). Nold (1980) also introduced the ideas of proportionate mixing and a convex combination of internal and proportionate mixing (named preferred mixing by Jacquez *et al.* 1988 and biased mixing by Hyman and Stanley 1988, 1989). Note, however, that different results have been obtained for an SIR model with variable population size for these two types of mixing by Castillo-Chavez *et al.* (1989c) and Huang *et al.* (1989). For further details see Castillo-Chavez *et al.* (1989d) in this volume.

The results of a long-term gonorrhea modeling project have been described in the book by Hethcote and Yorke (1984). The  $n$ -group model used was

$$\frac{dN_i I_i}{dt} = \left[ \sum_{j=1}^n \lambda_{ij} N_i I_i \right] (1 - I_i) - N_i I_i / d_i,$$

where  $N_i$  is the group  $i$  population size (assumed constant),  $\lambda_{ij}$  is the average number of adequate contacts per unit time of a person in group  $j$  with all people in group  $i$ , and  $d_i$  is the

average duration of infection in group  $i$ . The concept of a core group was introduced and control strategies were compared in a core-noncore model. Also, seasonal oscillations were analyzed in a female-male model. Six contact methods were compared in an eight-group model where people were divided by sex, sexual activity level, and whether infection was symptomatic or asymptomatic. Many of the multigroup models for HIV transmission and AIDS are similar in some ways to the gonorrhea model of Hethcote and Yorke (1984).

A multigroup model with both local and broader mixing was developed by Sattenspiel (1987) for the spread of Hepatitis A among day care centers in Albuquerque, New Mexico. This model included neighborhood play groups whose members mixed together by going to day care centers. This model was analyzed mathematically in Sattenspiel (1987) and Sattenspiel and Simon (1988).

Recently, multigroup models have been used for AIDS by several different authors. Instead of attempting to describe the chronological development of all of these models, the work of several authors or working groups is described. The selection is biased by our viewpoints and interests; no attempt is made to include all modeling efforts.

A working group at Los Alamos National Laboratory has analyzed models for HIV transmission and AIDS which include heterogeneity in sexual activity and mixing. In these models, which could be called continuous models, heterogeneity is introduced as continuous distributions of sexual behavior instead of as discrete groups with different behaviors. Hyman and Stanley (1988, 1989) formulated and used several models to study questions related to the AIDS epidemic. Their  $\tau$ -dependent model, where  $\tau$  denotes time since infection, includes variable infectivity as a function of  $\tau$ . This model is given by a system of nonlinear integro-differential equations for the distribution of infecteds and AIDS cases as a function of time and age since infection. Sample calculations showed that the infectivity profile could dramatically change the rate at which the susceptible population is infected. In their models, they used a Weibull distribution for the AIDS incubation period, and initial cubic growth of the AIDS cases and inverse quartic distributions for the number of sexual partners per unit time. They also used risk-based models with random (proportionate) mixing and biased (preferred) mixing. With random mixing, their numerical simulations showed that the disease progresses rapidly in both the high and low risk populations, but with biased (like-with-like) mixing, the disease progresses rapidly in the high risk populations and much more slowly in the low risk populations. The random mixing result seems inconsistent with data. They also noted that if the difference between the male-to-female and female-to-male infectivities is large, then the lower of these two infectivities tends to determine heterosexual spread. The number of infected people as a function of time can be determined by a convolution integral from the AIDS incidence as a function of time and the distribution of the AIDS incubation period. Back-calculation (discussed in Section 3.2) estimated the HIV incidence as a function

of time from the AIDS incidence and as an assumed distribution of the AIDS incubation period. Hyman and Stanley analyzed the accuracy of this procedure and estimated confidence intervals on forecasts based on the accuracy of AIDS data and the incubation time. Recently, Hyman and Stanley (1989) explored the sensitivity of an HIV transmission model to different social mixing patterns. This model was more general than that in Hyman and Stanley (1988) since it simultaneously included both continuous distributions for sexual partner change rates and also variable infectivity as a function of time since infection. A Weibull distribution was used for the AIDS incubation period. An acceptance (or preference) function determined which partners are acceptable to an individual and defined the mixing between groups with different partner change rates. Hyman and Stanley found that, if people select partners with very similar risk behavior, then the epidemic grows much more slowly than if they were more random in selecting partners. This reinforces the results in the previous paper.

Busenberg and Castillo-Chavez (1989b) have generalized this model to include age-structure. Furthermore, Busenberg and Castillo-Chavez (while generalizing the mixing framework of Blythe and Castillo-Chavez 1989) have found an explicit formula — the mixing function — that includes all mixing patterns for a homosexually active age-structured population as a function of an arbitrary preference (acceptance) function. Furthermore, they have determined an explicit formula for the reproductive number, in the case of proportionate mixing, for this very general model. A brief article discussing these results (which also generalize those of the Michigan group, described below) appears in this volume (see Busenberg and Castillo-Chavez 1989a).

Colgate *et al.* (1989) observed that the cumulative number of AIDS cases in the United States has grown as the cube of time rather than exponentially. They used a risk behavior model in which individuals tend to mix with people in their own risk group to explain that cubic growth is the result of a saturation wave of infection moving from the high to the low risk groups. Thus they claim that the observed decreasing growth rate in AIDS cases is due to the mixing pattern of the risk groups rather than to changes in behavior.

A primary focus of a working group centered at the University of Michigan has been to analyze the effects of various mixing patterns in discrete multigroup models for HIV and AIDS. In Jacquez *et al.* (1989), they presented a compartmental model for the spread of HIV in a homosexual population divided into subgroups by level of sexual activity. Their model included constant recruitment into the susceptible classes and variable infectivity in the stages of infection leading up to AIDS. They introduced the term “preferred mixing” for the situation in which a fraction of each group’s contacts are internal and the balance is distributed using proportionate mixing among all groups. This convex combination approach was not given a name by Nold (1980) and is similar to “biased mixing” in Los Alamos working group models. Both types of mixing are special cases of the like-with-like mixing of

Blythe and Castillo-Chavez (1989), and they are included in the representation theorem of Busenberg and Castillo-Chavez (1989a, b). The Michigan group determined a threshold for their model involving the mean number of contacts per infective, which determines whether or not there exists a unique endemic equilibrium (under some special conditions) in addition to the trivial (infection-free) equilibrium. For special cases, they were able to analyze the local stability of these equilibria. For a five-group model, numerical simulations revealed certain features. Increasing contacts between the groups greatly increased the rate of spread of the epidemic in the low activity level groups and also increased the fractions within those groups eventually infected. The conclusions of the Los Alamos working group and the Michigan working group are similar even though the former used continuously distributed models (and based its conclusions mostly on numerical simulations) and the latter used multigroup models and provided a local stability analysis of their model. They both found, through numerical simulations, that the mixing patterns between people with different sexual activity levels influenced the speed and asymptotic behavior of the epidemic.

A discrete  $n$ -group model that includes the Michigan model (Jacquez *et al.* 1988) has been analyzed recently by Castillo-Chavez *et al.* (1989c, d) and Huang *et al.* (1989) in the case of biased or preferred mixing and in the case of proportionate mixing. These researchers have found that the incorporation of a less restrictive incidence term allows for the possibility of two endemic equilibria. The bifurcation parameter is given by the average length of sexual activity, and non-symmetric mixing appears to play a very important role in the bifurcation two multiple equilibria. The article by Castillo-Chavez *et al.* (1989c) in this volume discusses these results in further detail, and provides two artificial examples, using proportionate and preferred mixing, for which asymmetry in epidemiological or mixing parameters leads to the existence of at least two endemic equilibria.

Jacquez *et al.* (1989, this volume) examined a model with mixing occurring in mixing subgroups that cut across "structural" subgroups. A mixing group matrix specifies the allocation of structural subgroups to the mixing subgroups. They also defined contact matrices for the mixing subgroups and gave examples. They called this "structured mixing".

Koopman *et al.* (1989, this volume) consider selective mixing which specifies the selection of contacts within a mixing group. The contact matrix involves conversation contact activity levels, the mutual acceptability of conversation contacts for sex, and the proportion of mutually acceptable encounters which result in new sexual partnerships. An example for a homosexual male population shows the difficulty of estimating the relative risk of HIV infection during anal and oral sex.

Knox (1986) used a multigroup model for transmission of AIDS. The twelve groups considered included male homosexuals divided into anal penetrative, anal receptive or both, and divided into promiscuous and not promiscuous. Equilibrium levels were predicted for the

groups, and the times necessary to reach these levels were determined. Because of the uncertainties of his plausible parameter set, he also analyzed forecasts using parameter sets from the surrounding parameter spaces.

In contrast to the many models for a homosexual male population, De Gruttola and Mayer (1987) considered a model for two interacting groups: a small population of individuals rapidly infected by high risk activity and a large population of individuals at risk only from heterosexual contact. The model was a system of differential equations in the male and female risk groups of intravenous drug users and their heterosexual partners. Individuals were either susceptible, asymptotically infected or symptomatically infected. Fitted parameter values were restricted to a range consistent with findings from partner studies. Their modeling showed that the interpretation of existing data on heterosexual transmission depends strongly on the infectiousness over time of HIV infecteds. Moreover, they showed that epidemics of widely varying severity among heterosexuals are consistent with the available epidemiological data.

Gail *et al.* (1989) developed a model for AIDS in order to evaluate the potential benefits of voluntary confidential screening for HIV. The model includes very active homosexuals, active homosexuals, bisexuals, heterosexual males, and heterosexual females. The contact matrix is determined by minimizing the discriminant information which is different from proportionate mixing. Conclusions are drawn about the usefulness of voluntary confidential screening in the high and low risk groups in a population of about 100,000 over a period of 5 to 15 years. The economic ratio (ER) is defined as the number of tests required to prevent one case. Voluntary confidential testing (VCT) prevents many cases in isolated high risk populations where ER values are less than 100. VCT prevents only a few cases in isolated low risk populations with ER values over 2,000 when the initial prevalence is below 0.1%, but prevents many more cases when the initial prevalence is above 1%. This paper quantifies the conclusion that screening is more effective in high risk populations than in low risk.

A general model for HIV transmission and AIDS has been formulated by Hethcote (1987, 1989a). The comprehensive model proposed contains all known transmission routes including homosexual and heterosexual intercourse, needle sharing among intravenous drug users, blood transfusions, blood factor concentrates to hemophiliacs, and perinatal infections. The primary risk groups in the model were sexually active homosexual and bisexual men, prostitutes, sexually active heterosexual women and men, and intravenous drug using women and men. The secondary risk groups were transfusion recipients, hemophiliacs, monogamous partners and children born to women in a previous risk group. For each risk group there was a differential equation incorporating the inflow and outflow. The progression from HIV infection to AIDS was modeled by a unidirectional flow in a sequence of stages. No attempt was made in Hethcote (1987, 1989a) to estimate parameter values or to apply the model.

In Hethcote (1989b), an HIV transmission and AIDS model was formulated as a system of nonlinear difference equations with a time step of one month. A priori estimates of the parameters were obtained from the literature. Parameters were estimated for group population sizes, migration, mortality, sexual activity levels and contact rates, needle sharing behavior, and state transition rates. Because of the uncertainty in some of these parameter estimates, more focused models have been used in specific geographic locations. Because both HIV prevalence and AIDS incidence data are available for homosexual males in San Francisco, Hethcote (1989c) focused on this population. After presenting the data, parameters related to the stages leading to AIDS and also to homosexual behavior were estimated. A best fit which optimized the fit criterion was obtained, and the sensitivity of the patterns to parameter changes was analyzed. The data and the model yielded a distribution of the AIDS incubation period which gave the best fit. A forecast based on the best fit was given. Hethcote (1989c) also concluded that the observed HIV prevalence and AIDS incidence in San Francisco were due to changes in sexual behavior and not due solely to saturation in high risk groups.

### 5.3 Demographic models

One of the most pressing issues in developing nations is that of determining what are the demographic consequences of the AIDS epidemic. Noticeable changes in the age-structure of a population can affect social and economic structure significantly even without observable reductions in population growth rates. Models help to outline possible scenarios, sort out the possibilities, reject or support hypotheses and identify specific data needs.

The history of epidemiological models that incorporate age-structure goes back to the influential paper of Hoppensteadt (1974). Hoppensteadt's article has generated a tremendous amount of activity in mathematical epidemiology and has influenced the field of theoretical biology. For some of the most recent applications see Castillo-Chavez *et al.* (1988, 1989e) and references therein.

Until recently, most of the demographic models of HIV transmission were developed and analyzed by May, Anderson, and their collaborators. May *et al.* (1988, 1989b) and Anderson *et al.* (1988c) introduced some of the first age-structured models for the sexual transmission of HIV. Conventional demographic models (see Hoppensteadt 1974, 1975; Castillo-Chavez 1989) were combined by May, Anderson and McLean with simple epidemiological models in order to explore the effects of horizontally- and vertically-transmitted AIDS upon total population growth rates as well as upon age-profile growth rates. The main objective of these papers was to use simple models to gain some qualitative understanding for reference in



numerical simulations of more realistic models. Model results suggest that AIDS is capable of changing population growth rates from positive to negative over a period of a few decades. Further results suggest that AIDS may have a minimal impact on the dependency ratio (the number of individuals below age 15 and above age 64 divided by the number of individuals in the complementary age classes).

The explosion of theoretical work on mixing (see Castillo-Chavez and Blythe 1989, Busenberg and Castillo-Chavez 1989a, and references therein) has demonstrated that further understanding of HIV transmission depends heavily on a better understanding of social/sexual mixing patterns. Progress is, therefore, limited by the lack of accurate data on sexual/social mixing patterns. The recent work of Gupta *et al.* (1989) and Anderson *et al.* (1989a) begins to look systematically at mixing patterns and their relationship to dynamic models. The representation theorem of Busenberg and Castillo-Chavez of all mixing patterns for a risk-based and age-structured homosexually-active population via a preference or acceptance function has provided a solid mathematical foundation upon which the effects of social/sexual mixing can be systematically analyzed.

Most of the models described above do not take into consideration the temporary periods of immunity that are naturally provided by short or long-term monogamous relationships. These effects can only be considered through models that follow the dynamics of pairs, i.e., models that explicitly take into consideration the processes of pair formation and pair dissolution. Demographic models that consider the age structure of a population and follow the dynamics of pairs were introduced by Fredrickson (1971). The first epidemiological model incorporating pairs was formulated in Dietz and Hadelar (1988) and Dietz (1988). Extensions of these models to age-structured populations were introduced and partially analyzed by Hadelar (1989a b). As stated earlier, the work of Dietz, Hadelar and collaborators suggests that the incorporation of the processes of pair formation and pair dissolution have a significant downward effect on the magnitude of the AIDS epidemic, when compared with the results of classical models.

Although these modeling approaches provide us with valuable insights, it is clear that they represent two extremes, and therefore the development and analysis of hybrid models is of extreme importance (see Waldstätter 1989). An alternative approach for the construction of hybrid two-sex models starts by postulating the existence of two mixing functions (one for females, the other for males) that satisfy appropriate mixing axioms. Solutions to these axioms — or mixing functions — can be incorporated into models with pairs and models without pairs, and possibly to hybrid models. Some specific cases are being studied by Castillo-Chavez and Busenberg (1989) but further work is needed.

## 6. Conclusion

This paper presents a partial review of the literature on mathematical and statistical approaches to AIDS epidemiology . Mathematical models for AIDS cannot be built in a vacuum. They have to be guided and motivated by data. Unfortunately, data is not readily available and there are many difficulties when dealing with the analysis of the available AIDS data (see Section 2). Models have been used quite successfully for short-term predictions of AIDS incidence and, with some degree of success, in estimating the number of infected individuals and the incubation period distribution. Dynamic models can be very useful in reducing the uncertainties in many of these estimates (see Section 3). Further, they allow us to see the effects that of a varying environment (due to changes in behaviors, medical treatment, etc.) on key parameters. Therefore, models can be irreplaceable in the evaluation of the relative merits of preventive strategies.

There is strong agreement that variable infectivity, the nature and type of the social/sexual mixing structures, and the long and variable period of infectiousness are key factors that must be understood if we expect to have some degree of success in the fight against this epidemic. Although there will always be the need for further mathematical studies, it is clear that in order to have an immediate impact in the fight against AIDS, we need to tie models to data and we need more quality data. Especially urgently required is data on mixing patterns and variable infectivity.

Mathematical epidemiology has shown a tremendous growth over the last two to three years. The challenging questions posed by the AIDS epidemic have focused the research of several scientists and have opened new directions of research, many of which are illustrated in this volume. Unfortunately, the scientific growth spurred by this epidemic may be one of the very few positive consequences of an otherwise frightful situation. This epidemic has created world-wide fear and pain, and it has just begun to affect significantly the economies of developed and developing nations through severe taxing of their medical services. At present it appears that this epidemic will eventually affect the lives of all individuals. We must keep this situation in mind as we continue to test and develop mathematical and statistical models for AIDS.

## ACKNOWLEDGMENTS

This research has been partially supported by NSF grant DMS-8906580, NIAID grant RO1 A129178-01, and Hatch project grant NYC 151-409, USDA awarded to C. C.-C. We thank S.A. Colgate, K.L. Cooke, C.M. Crawford, K. Dietz, J.M. Hyman, A. Levin, E.A. Stanley, and R. Waldstätter for their stimulating conversations.

## REFERENCES

- Anderson, R.M. (1988a). The epidemiology of HIV infection: variable incubation plus infectious periods and heterogeneity in sexual activity. *J. R. Statist. Soc. A*, 151, 66-93.
- Anderson, R.M. (1988b). The role of mathematical models in the study of HIV transmission and the epidemiology of AIDS. *J. AIDS* 1, 241-256.
- Anderson, R.M. (1989). Mathematical and statistical studies of the epidemiology of HIV. *AIDS* 3, 333-346.
- Anderson, R.M., S.P. Blythe, G.F. Medley, and A.M. Johnson. (1987). Is it possible to predict the minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United Kingdom? *Lancet* 1073-1075, May 9.
- Anderson, R.M., S.P. Blythe, S. Gupta, E. Konnings. (1989a). The transmission dynamics of the human immunodeficiency virus type 1 in the male homosexual community in the United Kingdom: the influence of changes in sexual behavior. (Manuscript.)
- Anderson, R.M. and R.M. May. (1989). Complex dynamical behaviour in the interaction between HIV and the immune systems. In *Cell to Cell signalling: from experiment to theoretical models*, Academic Press, 355-349.
- Anderson, R.M., R.M. May, A.R. McLean. (1988). Possible demographic impact of AIDS in developing countries. *Nature* 332, 228-234.
- Anderson, R.M., G.F. Medley, R.M. May, and A.M. Johnson. (1986). A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. *IMA J. of Mathematics Applied in Med. and Biol.* 3, 229-263.
- Aron, J.L. and P.S. Sarma. (1989). Assessment of the risk of HIV spread via non-steady heterosexual partners in the U.S. population. (This volume.)
- Bailey, N.T.J. (1989). The modeling and prediction of HIV/AIDS. (Manuscript.)
- Baldwin, J.D. and J.I. Baldwin. (1988). Factors affecting AIDS-related sexual risk-taking behavior among college students. *J. Sex Research* 25, 181-196.
- Banks H.T., and B.G. Fitzpatrick. (1989). Inverse problems for distributed systems: statistical tests and ANOVA. In *Mathematical approaches to problems in resource management and epidemiology*. C. Castillo-Chavez, S.A. Levin, and C. Shoemaker (eds.). Lecture Notes in Biomathematics 81, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo. (In press.)
- Bernoulli, D. (1760). Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir. *Mém. Math. Phys. Acad. Roy. Sci.*, Paris, pp. 1-45.

- Blythe, S.P. and R.M. Anderson. (1988a). Distributed incubation and infectious periods in models of transmission dynamics of human immunodeficiency virus (HIV). *IMA. J. of Mathematics Applied in Med. and Biol.* 5, 1-19.
- Blythe, S.P. and R.M. Anderson. (1988b). Variable infectiousness in HIV transmission models. *IMA J. of Mathematics Applied in Med. and Biol.* 5, 181-200.
- Blythe, S.P. and C. Castillo-Chavez. (1989). Like-with-like preference and sexual mixing models. *Math. Biosci.* 96, 221-238.
- Bradburn, N.M. (1983). *Response effects.* In: *Handbook of Survey Research*, P.H. Rossi, J.D. Wright, and A.B. Anderson (Eds.). Quantitative Studies in Social Relations, Academic Press, New York.
- Brookmeyer, R. and A. Damiano. (1989). Statistical methods for short term projections of AIDS incidence. *Statist. Med.* 8. (In press.)
- Brookmeyer, R. and M.H. Gail. (1986). Minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United States. *Lancet* 1320-1322, December 6.
- Brookmeyer, R. and M.H. Gail. (1988). A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *J. Amer. Stat. Assn.* 83, 301-308.
- Brookmeyer, R., M.H. Gail, and B.F. Polk. (1987). The prevalent cohort study and the acquired immunodeficiency syndrome. *Amer. J. Epidem.* 126, 14-24.
- Brownlee, J. (1907). Statistical studies in immunity. The theory of an epidemic. *Proc. Soc. Edinburgh* 26, 484-521.
- Busenberg, S. and C. Castillo-Chavez. (1989a). Interaction, pair formation and force of infection terms in sexually transmitted diseases. (This volume.)
- Busenberg, S. and C. Castillo-Chavez. (1989b). Risk and age-dependent mixing functions and force of infection terms in sexually transmitted diseases. (Submitted.)
- Cardell, N.S. and K.E. Kanouse. (1989). Population heterogeneity in a model of the spread of HIV infection. (This volume.)
- Castillo-Chavez, C. (1989). Some applications of structure models in population dynamics. In *Applied Mathematical Ecology*. S.A. Levin, T.G. Hallam, and L.J. Gross (eds.). *Biometrics* 18, 450-470. Springer-Verlag, Berlin, Heidelberg, New York, Tokyo. (In press.)
- Castillo-Chavez, C. and S.P. Blythe. (1989). Mixing framework for social/sexual behavior. (This volume.)
- Castillo-Chavez, C. and S. Busenberg. (1989). Pair formation in age- and risk-structured populations. (Manuscript in preparation.)

- Castillo-Chavez, C., K. Cooke, W. Huang, and S.A. Levin. (1989a). On the role of long periods of infectiousness in the dynamics of acquired immunodeficiency syndrome (AIDS). In *Mathematical approaches to problems in resource management and epidemiology*. C. Castillo-Chavez, S.A. Levin, and C. Shoemaker (eds.). Lecture Notes in Biomathematics 81, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo. (In press.)
- Castillo-Chavez, C., K. Cooke, W. Huang, and S.A. Levin. (1989b). On the role of long incubation periods in the dynamics of acquired immunodeficiency syndrome (AIDS), Part 1. Single population models. *J. Math. Biol.* 27, 373-398.
- Castillo-Chavez, C., K. Cooke, W. Huang, and S.A. Levin. (1989c). Results on the dynamics for models for the sexual transmission of the human immunodeficiency virus. *Applied Mathematics Letters*. (In press.)
- Castillo-Chavez, C., K. Cooke, W. Huang, and S.A. Levin. (1989d). On the role of long incubation periods in the dynamics of acquired immunodeficiency syndrome (AIDS), Part 2. Multiple group models. (This volume.)
- Centers for Disease Control. (1987). Revision of the CDC surveillance case definition for acquired immunodeficiency syndrome. *MMWR* 36, 3S-15.
- Colgate, S.A., E.A. Stanley, J.M. Hyman, S.P. Layne, and C. Qualls. (1989). A risk based model for explaining the cubic growth of AIDS cases. LA-UR-873412. (Los Alamos technical report.)
- Conover, W.J. (1980). *Practical Nonparametric Statistics*, 2nd edition. John Wiley & Sons, New York.
- Cooke, K.L. and J.A. Yorke. (1973). Some equations modeling growth processes and gonorrhea epidemics. *Math. Biosci.* 16, 75-101.
- Cox, D.R. (1972). Regression models and life tables. *J. R. Statist. Soc. B*, 34, 187-220.
- De Gruttola, V. and S.W. Lagakos. (1989a). Analysis of doubly-censored survival data, with application to AIDS. *Biometrics* 45, 1-11.
- De Gruttola, V. and S.W. Lagakos. (1989b). Epidemic models, empirical studies, and uncertainty. (This volume.)
- De Gruttola, V. and K.H. Meyer. (1987). Assessing and modeling heterosexual spread of human immunodeficiency virus in the United States. *Reviews of Infectious Diseases*, 10(1), 138-150.
- Dietz, K. (1988a). The first epidemic model: a historical note on P. D. E'nko, *Austral. J. Statist.* 30(A), 56-65.
- Dietz, K. (1988b). On the transmission dynamics of HIV. *Math. Biosci.* 90, 397-414.
- Dietz, K. and K.P. Haderler. (1988). Epidemiological models for sexually transmitted diseases. *J. Math. Biol.* 26, 1-25.

- Downs, A.M., R.A. Ancelle, H.J.C. Jager, and J-B. Brunet. (1987). AIDS in Europe: current trends and short-term predictions estimated from surveillance data, January 1981 - June 1986. *AIDS* 1, 53-57.
- Eubank, R. (1988). *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, Inc. New York.
- Fredrickson, A.G. (1971). A mathematical theory of age structure in sexual populations: random mating and monogamous marriage models. *Math. Biosci.* 10, 117-143.
- Fox, J.A. and P. Tracy. (1984). *Randomized Response*. Sage, Beverly Hills, CA.
- Fusaro, R.E., N.P. Jewell, W.W. Hauck, D.C. Heilbron, J.D. Kalbfleisch, J.M. Neuhaus, and M.A. Ashby. (1989). An annotated bibliography of quantitative methodology relating to the AIDS epidemic. *Statistical Science* 4, 264-281.
- Gail, M.H., D. Preston, and S. Piantadosi. (1989). Disease prevention models of voluntary confidential screening for human immunodeficiency virus (HIV) in isolated low risk and high risk populations and in mixed gay/heterosexual populations. *Statistics in Med.* 8. (In press.)
- Giesecke, J., G. Scalia-Tomba, O. Berglund, E. Berntorp, S. Schulman, and L. Stigendal. (1988). Incidence of symptoms and AIDS in 146 Swedish hemophiliacs and blood transfusion recipients infected with human immunodeficiency virus. *Br. Med. J.* 297, 99-102.
- Gilchrist, W. (1983). Forecasting. In *Encyclopedia of Statistical Sciences*. S. Kotz, N.L. Johnson, and C.B. Read (eds.). Volume 3, John Wiley & Sons, New York.
- Goedert, J.J., R.J. Biggar, M. Melbye, D.L. Mann, S. Wilson, M.H. Gail, R.J. Grossman, R. A. DiGioia, W.C. Sanchez, S.H. Weiss, and W.A. Blattner. (1987). Effect of T4 count and cofactors on the incidence of AIDS in homosexual men infected with human immunodeficiency virus. *J. Amer. Med. Assn.* 257, 331-334.
- Gupta, S., R.M. Anderson, and R.M. May. (1989). Networks of sexual contacts: implications for the pattern of spread of HIV. (Manuscript.)
- Hadeler, K.P. (1987). Pair formation in age structured populations. *Proceedings, Workshop on Selected Topics in Biomathematics*, IIASA. Kurzhanshij, A. and K. Sigmund (eds.). Laxenburg, Austria.
- Hadeler, K.P. (1989). Pair formation in age-structured populations. *Acta Applicandae Mathematicae* 14, 91-102.
- Hadeler, K.P., R. Waldstätter, and A. Wörz-Busekros. (1988). A model for pair formation in bisexual populations. *J. Math. Biol.* 26, 635-649.
- Harris, J.E. (1987). Delay in reporting acquired immune deficiency syndrome (AIDS). Working paper 452, Department of Economics, MIT. Cambridge, Mass.

- Healy, M.J.R. and H.E. Tillett. (1988). Short-term extrapolation of the AIDS epidemic. *J. R. Statist. Soc. A*, 151, 50-61.
- Hethcote, H.W. (1976). Qualitative analyses for communicable disease models, *Math. Biosci.* 28, 335-356.
- Hethcote, H.W. (1978). An immunization model for a heterogeneous population. *Theor. Pop. Biol.* 14, 338-349.
- Hethcote, H.W. (1987). AIDS modeling work in the USA. In *Future Trends in AIDS*. Her Majesty's Stationery Office, London, 35-40.
- Hethcote, H.W. (1989a). A model for HIV transmission and AIDS. In *Mathematical approaches to problems in resource management and epidemiology*. C. Castillo-Chavez, S. A. Levin, and C. Shoemaker (eds.). Lecture Notes in Biomathematics 81, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo. (In press.)
- Hethcote, H.W. (1989b). Model formulation and parameter estimation, phase 1 progress report to CDC. (Unpublished report.)
- Hethcote, H.W. (1989c). HIV prevalence and AIDS in San Francisco, phase 2 progress report to CDC. (Unpublished report.)
- Hethcote, H.W. and H.R. Thieme. (1985). Stability of the endemic equilibrium in epidemic models with subpopulations. *Math. Biosci.* 75, 205-227.
- Hethcote, H.W. and J.A. Van Ark. (1987). Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation and immunization programs. *Math. Biosci.* 84, 85-118.
- Hethcote, H.W. and J.A. Yorke. (1984). *Gonorrhea transmission dynamics and control*. Lecture Notes in Biomathematics 56, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo.
- Hoffman, G.W. and M.D. Grant. (1989). When HIV meets the immune system: network theory, alloimmunity and AIDS. (This volume.)
- Hoppensteadt, F. (1974). An age dependent epidemic model. *J. Franklin Instit.* 297, 325-333.
- Hoppensteadt, F. (1975). Mathematical Theories of Populations: demographics, genetics and epidemics. *SIAM Regional Conference Series in Applied Math.*, No. 20. Philadelphia.
- Huang, W., K. Cooke, and C. Castillo-Chavez. (1989). Stability and bifurcation for a multiple group model for the dynamics of HIV/AIDS. (Submitted.)
- Hyman, J.M. and E.A. Stanley. (1988). Using mathematical models to understand the AIDS epidemic. *Math. Biosci.* 90, 415-473.

- Hyman, J.M. and E.A. Stanley. (1989). The effect of social mixing patterns on the spread of AIDS. In *Mathematical approaches to problems in resource management and epidemiology*. C. Castillo-Chavez, S.A. Levin, and C. Shoemaker (eds.). Lecture Notes in Biomathematics 81, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo. (In press.)
- Isham, V. (1988). Mathematical modeling of the transmission dynamics of HIV infection and AIDS: a review. *J. R. Statist. Soc. A*, 151, 5-30.
- Jacquez, J.A., C.P. Simon, and J. Koopman. (1989). Structured mixing: heterogeneous mixing by the definition of mixing groups. (This volume.)
- Johnson, A.M. (1988). Social and behavioural aspects of the HIV epidemic - a review. *J. R. Statist. Soc. A*, 151, 99-114.
- Kalbfleisch, J.D. and J.L. Lawless. (1989). Inference based on retrospective ascertainment: an analysis of the data on transfusion-related AIDS. *J. Amer. Stat. Assn.* 84, 360-372.
- Kaplan, E.H. (1989). Needles that kill: modeling human immunodeficiency virus transmission via shared drug injection equipment in shooting galleries. *Rev. of Inf. Diseases*. (In press.)
- Kaplan, E.H., P.C. Cramton, and A.D. Paltiel. (1989). Nonrandom mixing models of HIV transmission. (This volume.)
- Kaplan, H.B., R.J. Johnson, C.A. Bailey, and W. Simon. (1987). The sociological study of AIDS: a critical review of the literature and suggested research agenda. *J. Health and Social Behavior* 28, 140-157.
- Karon, J.M., O.J. Devine, and W.M. Morgan. (1989). Predicting AIDS incidence by extrapolating from recent trends. (This volume.)
- Karon, J.M., T.J. Dondero, Jr., and J.W. Curran. (1988). The projected incidence of AIDS and estimated prevalence of HIV infection in the United States. *J. AIDS* 1, 542-550.
- Kendall, D.G. (1949). Stochastic processes and population growth. *Roy. Statist. Soc., Ser. B* 2, 230-264.
- Kermack, W.O. and A.G. McKendrick. (1927). A contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. London, Ser. A* 115, 700-721.
- Knox, E.G. (1986). A transmission model for AIDS. *Eur. J. Epidemiol.* 2, 165-177.
- Koopman, J.S., C.P. Simon, and J.A. Jacquez. (1989). Selective contact within structured mixing groups; with an application to the analysis of HIV transmission risk from oral and anal sex. (This volume.)
- Lagakos, S.W., L.M. Barraj, and V De Gruttola. (1988). Nonparametric analysis of truncated survival data, with application to AIDS. *Biometrika* 75, 515-523.
- Lajmanovich, A. and J.A. Yorke. (1976). A deterministic model for gonorrhea in a nonhomogeneous population. *Math. Biosci.* 28, 221-236.



- Layne, S.P., T.G. Marr, S.A. Colgate, J.M. Hyman, and E.A. Stanley. (1988). The need for national HIV databases. *Nature* 333, 511-512.
- Layne, S.P., J.L. Sponge, and M. Dembo. (1989). Measuring HIV infectivity. (This volume.)
- Lehmann, E. L. (1975). *Nonparametrics*. Holden-Day, San Francisco.
- Lehmann, E. L. (1983). *Theory of Point Estimation*. John Wiley & Sons, New York.
- Lehmann, E.L. (1986). *Testing Statistical Hypotheses*, 2nd edition. John Wiley & Sons, New York.
- Longini, I.M., Jr., W.S. Clark, M. Haber, and C.R. Horsburgh, Jr. (1989). The stages of HIV infection: waiting times and infectious contact rates. (This volume.)
- Lui, K-J., W.W. Darrow, and G.W. Rutherford III. (1988). A model based estimate of the mean incubation period for AIDS in homosexual men. *Science* 240, 1333-1335.
- Lui, K.J., D.N. Lawrence, W.M. Morgan, T.A. Peterman, H.W. Haverkos, and D.J. Bragman. (1986). A model-based approach for estimating the mean incubation period of transfusion-associated acquired immunodeficiency syndrome. *Proc. National Acad. Science, USA* 83, 3051-3055.
- Lyter, D.W., R.O. Valdiserri, L.A. Kingsley, W.P. Amoroso, and C.R. Rinaldo. (1987). The HIV antibody test. Why gay and bisexual men want or do not want to know their results? *Public Health Rep* 102, 468-474.
- MacDonald, N. (1957). *The Epidemiology and Control of Malaria*. Oxford University Press, London.
- May, R.M. and R.M. Anderson. (1987). Transmission dynamics of HIV infection. *Nature* 326, 137-142.
- May, R.M. and R.M. Anderson, and A.R. McLean. (1988). Possible demographic consequences of HIV/AIDS: I. Assuming HIV infection always leads to AIDS. *Math. Biosci.* 90, 475-506.
- May, R.M. and R.M. Anderson, and A.R. McLean. (1989). Possible demographic consequences of HIV/AIDS: II. Assuming HIV infection does not necessarily leads to AIDS. In *Mathematical approaches to problems in resource management and epidemiology*. C. Castillo-Chavez, S.A. Levin, and C. Shoemaker (eds.). Lecture Notes in Biomathematics 81, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo. (In press.)
- McCusker, J., A.M. Stoddard, K.H. Mayer, J. Zapka, C. Morrison, and S.P. Saltzman. (1988). Effects of HIV antibody test knowledge on subsequent sexual behaviors in a cohort of homosexually active men. *Amer. J. Public Health* 78, 462-467.
- McKendrick, A.G. (1912). On certain mathematical aspects of malaria. In *Proc. Imperial Malaria Com.* pp. 54-66.

- Medley, G.F., R.M. Anderson, D.R. Cox, and L. Billard. (1987). Incubation period of AIDS in patients infected via blood transfusion. *Nature* 328, 719-721.
- Medley, G.F., L. Billard, D.R. Cox, and R.M. Anderson. (1988). The distribution of the incubation period for the acquired immunodeficiency syndrome (AIDS). *Proc. R. Soc. Lond. B*, 233, 367-377.
- Merrill, S.J. (1989). Modeling the interaction of HIV with cells of the immune response. (This volume.)
- Mode, C.J., H.E. Gollwitzer, and W. Herrman. (1989). A methodological study of a stochastic model of an AIDS epidemic. *Math. Biosci.* 92, 201-229.
- Mood, A.M., F.A. Graybill, and D.C. Boes. (1974). *Introduction to the Theory of Statistics*, 3rd edition. McGraw-Hill Book Company, New York.
- Morgan, W.M. and J.W. Curran. (1986). Acquired immunodeficiency syndrome: current and future trends. *Public Health Rep* 101, 459-465.
- Neter, J., W. Wasserman, and M.H. Kutner. (1985). *Applied Linear Statistical Models*, 2nd edition. Richard D. Irwin, Inc., Homewood, IL.
- Nold, A. (1980). Heterogeneity in disease-transmission modeling. *Math. Biosci.* 52, 227-240.
- Padian, N., L. Marquis, D.P. Francis, R.E. Anderson, G.W. Rutherford, P.M. O'Malley, and W. Winkelstein. (1987). Male-to-female transmission of human immunodeficiency virus. *J. Amer. Med. Assn.* 258, 788-790.
- Perelson, A.S. (1989). Modeling the interaction of the immune system with HIV. (This volume.)
- Pickering, J., J.A. Wiley, N.S. Padian, L.E. Lieb, D.F. Echenberg, and J. Walker. (1986). Modeling the incidence of acquired immunodeficiency syndrome (AIDS) in San Francisco, Los Angeles and New York. *Mathematical Modeling* 7, 661-698.
- Ross, R. (1911). *The prevention of malaria*, 2nd edition. John Murray, London.
- Royal Statistical Society. (1988). Statistical requirements of the AIDS epidemic. *J. R. Statist. Soc. A* 151, 127-130.
- Sattenspiel, L. (1987). Population structure and the spread of disease. *Human Biol.* 59, 411-438.
- Sattenspiel, L. and C.P. Simon. (1988). The spread and persistence of infectious diseases in structured populations. *Math. Biosci.* 90, 341-366.
- Schuman, H. and S. Presser. (1981). *Questions and Answers in Attitude Surveys*. Quantitative Studies in Social Relations, Academic Press, Inc., New York.
- Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York.
- Tan, W. Y. (1989). Some stochastic models of AIDS spread. *Stat. in Med.* 8, 121-136.

- Thieme, H.R. and C. Castillo-Chavez. (1989a). On the role of variable infectivity in the dynamics of the human immunodeficiency virus. (This volume.)
- Thieme, H.R. and C. Castillo-Chavez. (1989b). On the possible effects of infection-age-dependent infectivity in the dynamics of HIV/AIDS. (Manuscript.)
- Waldstätter, R. (1989). Pair formation in sexually-transmitted diseases. (This volume.)
- Working Group. (1988). *Short-term prediction of HIV infection and AIDS in England and Wales*. Her Majesty's Stationery Office, London.
- Workshop. (1988). *A National Effort to Model AIDS Epidemiology*. Office of Science and Technology Policy, Executive Office of the President, Washington, D.C.
- Yorke, J.A., H.W. Hethcote, and A. Nold. (1978). Dynamics and control of the transmission of gonorrhea. *Sexually Transmitted Diseases* 10, 72-76.